UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
BOOKSTACKS

CENTRAL CIRCULATION BOOKSTACKS

The person charging this material is responsible for its renewal or its return to the library from which it was borrowed on or before the **Latest Date** stamped below. **You may be charged a minimum fee of $75.00 for each lost book.**

Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.

TO RENEW CALL TELEPHONE CENTER, 333-8400

UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

AUG 0 1 1997

When renewing by phone, write new due date below previous due date. L162

Digitized by the Internet Archive
in 2012 with funding from
University of Illinois Urbana-Champaign

http://www.archive.org/details/theoryofoptimalb92100kras

FACULTY WORKING PAPER 92-0100

#20-157

330 STX
B385
1992:100 COPY 2

A Theory of Optimal Bank Size

The Library of the
FEB 17 1992
University of Illinois
of Urbana-Champaign

Stefan Krasa
Department of Economics
University of Illinois

Anne P. Villamil
Department of Economics
University of Illinois

Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 92-0100

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

January 1992

A Theory of Optimal Bank Size

Stefan Krasa
Anne P. Villamil

Department of Economics

A Theory of Optimal Bank Size

Stefan Krasa Anne P. Villamil*

First Draft: May 1991 This Draft: December 1991

Abstract

This paper provides a theory of optimal bank size determination with implications for the size distribution of banks in a model with asymmetric information and costly state verification. Production is subject to minimum scale requirements and two types of risk: diversifiable idiosyncratic project risk, and imperfectly diversifiable aggregate "macroeconomic" risk. We first show that delegated monitoring with two-sided simple debt contracts dominates direct investment if the cost of monitoring the intermediary is bounded and if the variance of the non-diversifiable macroeconomic risk is sufficiently small. We next show that: (i) banks are of finite size; (ii) bank size is inversely related to the bank's exposure to macroeconomic risk, and (iii) multiple banks co-exist with the same size within a locale but with (possibly) different sizes across locales.

*Address of the authors: Department of Economics, University of Illinois, 1206 South Sixth Street, Champaign, IL 61820.

We gratefully acknowledge useful comments from Anthony Courakis and financial support from the National Science Foundation (SES 89-09242).

1 Introduction

Recent research has studied how the structure of the financial system affects the transmission of business cycle shocks in an economy. See Gertler (1988) for an excellent survey of this emerging literature. An equally important but quite different problem is the following: How do business cycle fluctuations affect the structure of the financial system? This problem is important because all developed countries are subject to regular and recurrent business cycle fluctuations. Further, banks in most countries operate under restrictive regulations which limit their ability to insure fully against macroeconomic risk. For example, in the U.S. there are branching restrictions which limit the geographic operation of banks and portfolio restrictions which limit the types of assets that banks may hold (e.g., Savings and Loan Institutions have been subject to such restrictions). It is obvious that these restrictions can lead to non-diversifiable portfolio risk, but little is known about the implications for financial structure. This problem is of particular interest at the present time. The European Community is currently involved in a transition toward an economic and monetary union (EMU). To date, most discussions of the EMU have focused on the creation of a European central bank and its supervision of a single currency. However, once an administrative structure is in place, it will undoubtedly impose "pan European" restrictions. What are the likely implications of such restrictions for the future European banking system? This paper proposes a theoretical model that can be used to analyze this important policy question.

We propose a theory of optimal bank size determination which has implications for the size distribution of banks. We consider a costly state verification model with finitely many borrowers and lenders where production is subject to: (i) minimum project scale requirements, (ii) diversifiable idiosyncratic project risk, and (iii) a non-diversifiable aggregate (i.e., macroeconomic) risk. Agents can undertake production in two ways. Borrowers and lenders can write direct bilateral investment contracts, or they can engage in intermediated investment by contracting with a bank that accepts deposits from lenders and grants loans to borrowers. Because there is non-trivial default risk in the economy, the lenders must monitor either the borrowers (in the direct investment problem) or the bank (in the intermediated investment problem) in default states which occur with strictly positive probability.

The minimum project scale requirement implies that it takes multiple

lenders to finance the project of a single borrower. It is well known (cf., Diamond (1984) or Williamson (1986)) that "delegated monitoring" may be optimal under this requirement because it allows lenders to economize on monitoring costs. However, in an economy with non-trivial default risk lenders must "monitor the monitor" (i.e., bank) because the bank may misreport the state to minimize its payments to lenders. Thus, the essential problem that lenders face is to provide the bank with an incentive to report truthfully to them. Krasa and Villamil (1991a) study this problem in a finite economy that is subject to *diversifiable* default risk. They show that a particular type of contract commonly issued by banks (i.e., two-sided simple debt) solves this monitoring problem optimally. In contrast, in this paper we are concerned with the optimal investment arrangement (given two-sided simple debt contracts) when there is both diversifiable project risk and *non-diversifiable* aggregate or "macroeconomic" risk. When a bank is subject to non-diversifiable project (and hence default) risk, costly monitoring will necessarily occur in some states—regardless of the bank's size.

In choosing an optimal portfolio size (i.e., scale of operation) a bank faces the following tradeoff for most monitoring cost structures. Increasing the size of the bank's portfolio (i.e., contracting with additional borrowers) given some initial bank size generally decreases the bank's default probability, *but* increases the lenders' cost of monitoring the bank. Thus, the crucial question that the bank faces in choosing an optimal portfolio size is: Under what circumstances do the gains from decreased default risk dominate the losses from increased monitoring costs when the bank compares its current scale of operation with an increased scale of operation? We begin our analysis of this question with Theorem 1 which shows that delegated monitoring with two-sided simple debt contracts (i.e., intermediated investment) dominates direct investment if the lenders' cost of monitoring the intermediary is bounded and the variance of the non-diversifiable macroeconomic risk is sufficiently small. We interpret the variance of the macroeconomic risk as the magnitude of business cycle fluctuations. For the U.S., Prescott (1986) reports that the standard deviation of output for the period 1872 to 1985 is only 1.8 percent.

Theorem 2 provides the main result of the paper: A theory of optimal bank size with implications for the size distribution of banks. To our knowledge, this has been a neglected branch of the literature on financial intermediation. Of course, the problem of firm size distribution has been studied extensively in industrial organization theory. Panzar (1989, p. 33) summa-

rizes the findings from this literature by noting that firm size "is determined in large part by the ... cost function," while industry structure (i.e., limits on the number and size distribution of firms that are present in equilibrium) "is determined by the market demand curve." Our Theorem 1 implies that this traditional industrial organization analysis is inadequate for a theory of financial structure. In Theorem 2 we develop a measure of the rate of portfolio diversification accruing from increases in bank size, and then show that both risk and cost considerations are essential determinants of bank size. The theory has three predictions that in principle are empirically testable. First, banks will be of finite size with the precise scale dependent upon the structure of monitoring costs *and* the degree of portfolio diversification that the bank can attain. Second, banks that are better able to diversify risk (e.g., because they are subject to less stringent portfolio restrictions) will be larger in size than banks which are less able to diversify risk. Third, multiple banks with similar risk and cost characteristics may co-exist. The first prediction pertains to firm size, while the latter two are industry structure predictions. Clearly, firm size and industry structure are related; however, we shall return to a more precise discussion of this relationship in Section 6.

2 The Model

Consider an economy with finite numbers of two types of risk-neutral agents, borrowers and lenders. Each borrower $i = 1, \ldots, \bar{n}$ is endowed with a risky investment project which transforms one unit of a single input at time zero into x_i units of output at time one, where x_i is the realization of a random variable X_i on the probability space $(\Omega, \mathcal{A}, P)$.[1] For simplicity assume that borrowers have zero endowment of the input. Every lender $j = 1, \ldots, \bar{m}$ is endowed with $a < 1$ units of a homogeneous input,[2] but has no direct access to a productive technology. Thus, the project of a borrower cannot be financed by a single lender which implies that in the absence of intermediation more than one lender would have to verify a single borrower. The total available supply of investment is larger than the input required by all borrowers, so $\bar{m}$ lenders can be accommodated by the $\bar{n}$ borrowers (i.e., $\bar{m}a \geq \bar{n}$).

[1] We will implicitly refer to this probability space when writing P for probability and E for expected value.

[2] For technical purposes assume that $1/a$ is an integer.

Also, assume there is a riskless alternative investment project available to all lenders that yields return r with probability one.

All borrowers and lenders are fully informed about the distribution of X_i at time zero, but asymmetric information exists about the state of the project's actual realization ex-post: Only borrower i costlessly observes the realization x_i of his/her project at time one. Let $F_i(x)$ denote the distribution of borrower i's project and assume that the F_i are identical (i.e., all X_i have the same distribution).[3] We now depart from the standard intermediation framework by considering an economy with non-trivial correlation among the X_i.[4] Assume that each X_i can be decomposed into independent random variables Y_i, $i = 1 \ldots, \bar{n}$, and Z, where Y_i is an idiosyncratic risk associated with borrower i's project and Z is a non-diversifiable "macroeconomic" risk common to all borrowers. Thus,

$$X_i = Y_i + Z,^{5} \tag{1}$$

Assume that the distributions of Y_i and Z have continuous density functions, $X_i \geq 0$ for every i (because borrowers can never produce "negative output" no matter how bad the macroeconomic shock), and that each X_i is bounded from above.

Let a technology exist which can be used by agents other than borrower i to verify at time one the realization x_i of project X_i. Assume that this state verification technology is costly to use, and that when verification occurs, x_i is privately revealed only to the individual who requests (deterministic) verification. Assume that the verification cost is comprised of both a pecuniary component and an indirect "pecuniary equivalent" of a non-pecuniary cost.[6] These costs may be thought of as the money paid to an attorney to file a claim (a pecuniary cost) and the monetary value of time lost when

[3]This assumption simplifies the analysis but is not essential for the results.

[4]See Dowd (1991) for an excellent survey of the literature on financial intermediation.

[5]To simplify the analysis we make the following technical assumption: Let $\Omega = \Omega_1 \times \Omega_2$ and let $P = P_1 \times P_2$, where P_i is a probability on Ω_i for $i = 1, 2$. Assume that the random variables Y_i are independent of Ω_2, i.e., for every $\omega_1 \in \Omega_1$ the mapping $\omega_2 \mapsto Y_i(\omega_1, \omega_2)$ is constant on Ω_2. Similarly, assume that $\omega_1 \mapsto Z(\omega_1, \omega_2)$ is constant on Ω_1. This condition implies independence of Z and X_i for every $i \in I\!N$, but is stronger than independence. In our analysis, standard independence would require conditions on Ω which imply the existence of a regular conditional probability $P(\cdot \mid Z)$ (cf., Parthasarathy (1977, Proposition 46.5)).

[6]The non-pecuniary costs permit negative utility but rule out negative consumption,

visiting the attorney (a pecuniary equivalent). Because agents have asymmetric information, a key problem is to ensure that the borrower reports the realization truthfully. We, like Williamson (1986), use the costly state verification framework to solve this problem. This model was introduced by Townsend (1979). However, unlike in Townsend's model where x_i is *publicly* announced after verification occurs, in our model x_i is *privately* revealed only to the agent who requests verification.[7] This assumption is essential for our analysis since if all information could be made public ex-post, there would be no need to verify the bank in default states. However, it also appears to accurately describe the privacy and institutional features which characterize most lending arrangements. For example, Diamond (1984, p. 395) observes: "Financial intermediaries in the world monitor much information about their borrowers in enforcing loan covenants, but typically do not directly announce this information or serve an auditor's function."

3 Contract Arrangements

There are two types of basic investment arrangements, notably those involving direct contracts between "primary" borrowers and lenders, and those involving an intermediary. In Section 3.1 we consider the direct investment problem. In Section 3.2 we consider intermediated investment where lenders and borrowers write contracts with a bank, and the bank is subject to non-trivial default risk.

3.1 Direct Investment

Let all direct, bilateral interactions between lenders and borrowers be regulated by a contract whose general form is defined as follows.

Definition 1. *A one-sided contract between lenders and borrowers is a pair* $(R(\cdot), S)$*, where* $R(\cdot)$ *is an integrable, positive payment function on* $\mathbb{R}_+$*,*

while pecuniary equivalents of non-pecuniary costs ensure that the costs can be shared by the contracting parties.

[7]See Krasa and Villamil (1991b) for an analysis of an economy with multiple heterogeneous agents, costly state verification, *public* announcement, and deterministic or stochastic verification.

such that $R(x) \leq x$ *for every* $x \in \mathbb{R}_+$ *and* S *is an open subset of* $\mathbb{R}_+$ *which determines the states where monitoring occurs.*

The contract $(R(\cdot), S)$ describes the total claims against the borrower by all lenders. If a lender invests b units of capital in a borrower's project, then his/her claim against the borrower is given by $bR(x)$, where x is the borrower's announced wealth realization. Following standard practice in this literature, we restrict the universe of contracts to the set of incentive-compatible contracts and denote this set by $C = (R(\cdot), S)$. Consequently, the realization announced by each borrower is the true realization. The following condition ensures that all contracts under consideration satisfy this restriction: There exists $\bar{R} \in \mathbb{R}_+$ such that $S = \{x: R(x) < \bar{R}\}$. The imposition of this restriction is without loss of generality because the Revelation Principle establishes that any arbitrary contract can be replaced by an incentive-compatible contract with the same actual payoff (cf., Townsend (1988, p. 416)). Therefore, the set of all incentive-compatible contracts is fully specified by the tuple $(R(\cdot), \bar{R})$.

We study a particular type of contract, called a simple debt contract, which is defined as follows:

Definition 2. $(R(\cdot), \bar{R})$ is a simple debt contract if: $R(x) = x$ for $x \in S = \{x < \bar{R}\}$ and $R(x) = \bar{R}$ if $x \in S^c = \{x \geq \bar{R}\}$.

The payment schedules in Definition 2 resemble simple debt because:

(i) When verification occurs the payment to the lender is state contingent (i.e., the borrower pays the entire realization for all outcomes below a cutoff level), where the verification set S is viewed as the set of bankruptcy states.

(ii) When verification does not occur the payment to the lender is constant (i.e., the borrower pays a fixed amount $\bar{R}$ for all realizations of the state above the cutoff), where S^c is the set of all realization where verification does not occur.

Townsend (1979) proved that debt contracts are optimal responses to asymmetric information problems in economies with deterministic costly state verification technologies because such contracts minimize verification costs. Agents verify only low realizations of X_i and accept fixed payments (which do not require monitoring) in all other states. Gale and Hellwig (1985)

and Williamson (1986) showed that *simple* debt is the optimal contract among all one-sided investment schemes. In contrast to debt contracts, where the borrower is the residual claimant in the default state (and hence may receive a non-zero payment if the bankruptcy is not "too severe"), a simple debt contract requires the borrower's *entire* project realization to be transferred to the lender in default states (i.e., see (i) above). This result will be useful in the analysis that follows, thus we state it formally.

Theorem GHW. *Simple debt is the optimal contract among all one-sided investment schemes.*

The strategy of the proof of Theorem GHW is as follows.[8] Consider two optimal contracts. Let $(R(\cdot), \bar{R})$ be a simple debt contract and $(A(\cdot), \bar{A})$ be some alternative contract. Since both contracts are optimal, both must provide borrowers with the same expected payoff. Under the simple debt contract lenders request costly state verification if $x < \bar{R}$, and under the alternative contract lenders request verification if $x < \bar{A}$. Clearly $\bar{A} > \bar{R}$ (otherwise the contracts cannot have the same expected return to borrowers), thus the expected verification costs must be less for the simple debt contract.

In our economy with correlation among project realizations, the direct investment problem is identical to the investment problem in an economy without correlation among projects. This follows from the fact that when borrowers and lenders write direct bilateral contracts, every lender must verify the borrower with whom he/she contracts in default states. Hence, nontrivial correlation among projects is irrelevant (and simple debt contracts remain optimal). Note that even though lenders have the opportunity to invest in more than one project (i.e., contract with more than one borrower), it is not optimal for them to do so because they reap no gain from diversifying idiosyncratic risk while they incur higher expected monitoring costs. For example, suppose that an agent invests in two projects. Assume that the total outstanding debt of borrowers $i = 1, 2$ is given by the contract $(R_i(\cdot), \bar{R}_i)$, and let a_i be the capital the lender invests in each project. If both projects do not fail, then the payoff is simply given by $a_1\bar{R}_1 + a_2\bar{R}_2$. Thus, there is no gain in the good state from investing in multiple projects. However, the probability that at least one of the two projects fail is strictly higher than the probability that only a single project fails.[9]

[8] See Gale and Hellwig (1985) or Williamson (1986) for a formal proof.

[9] For example, consider the random event to be the toss of a fair coin, where "head"

The direct investment problem between a borrower and lenders can now be stated. Let c denote the lenders' cost of monitoring a borrower.

Problem 3.1. Choose an incentive-compatible contract $(R(\cdot), S)$ to:

$$\max \int_0^T [x - R(x)]\, dF(x)$$

subject to:

$$a \int_0^T R(x)\, dF(x) - \int_S c\, dF(x) \geq ra. \tag{2}$$

In Problem 3.1 (the direct investment problem), the expected utility of a representative borrower is maximized subject to a constraint that the lenders' expected return, net of monitoring costs (c), be at least as great as some reservation level (r). The first term in the lenders' constraint is multiplied by a in order to account for each individual lender's capital investment a. Without loss of generality we assume that each lender invests all of his/her endowment in a single project. Finally, Problem 3.1 reflects the assumption that credit markets are competitive. There are more lenders who wish to invest than investment opportunities. Thus, the supply of loans is inelastic, and the level of return necessary to attract lenders is driven down to the reservation level r, the return available on the alternative investment opportunity.

3.2 Intermediated Investment

Now consider an intermediated borrowing and lending problem. In the previous section (i.e., the one-sided problem), lenders and borrowers wrote direct bilateral contracts and correlation among projects (as long as it was not trivial) was irrelevant. However, duplicative monitoring is inherent in the

is non-default, and "tail" is a default. Clearly, for a single coin toss the probability of a default is 0.5. If the agent "invests" in two coin tosses, the probability that at least one of the two projects fails is 0.75. Thus expected monitoring costs will be higher. This is true even if there is some correlation among projects, if idiosyncratic risk is non-trivial. If the idiosyncratic risk is trivial (i.e., $X_i = 0$ so only macroeconomic risk matters), then every agent could monitor only a single project (since the realization of all projects can be determined by the outcome of any one project) and the expected payoff from investing in one project or many projects is the same.

direct investment problem because each lender must verify each borrower with whom he/she contracts in certain states of nature. Thus, there may exist gains from "delegated monitoring" (cf., Diamond (1984)), where lenders elect a monitor to perform the verification task and thereby eliminate some of the duplicative monitoring associated with direct investment. In contrast to previous delegated monitoring studies, our economy has an important feature which significantly complicates the "standard" delegated monitoring problem. The intermediary faces non-trivial default risk for two reasons:

(i) Since there are only finitely many borrowers, it is not clear that the intermediary can completely diversify idiosyncratic risk.

(ii) Even if the intermediary can eliminate idiosyncratic risk, its portfolio is still subject to non-diversifiable "macroeconomic" risk.

Thus, in our economy the probability that the bank may default is non-zero (because at least the macroeconomic risk is non-diversifiable), so lenders must verify the bank with strictly positive probability (i.e., in some states).

We begin our analysis of the delegated monitoring problem by considering how agents select an intermediary. Since the loan market is competitive, any lender who wishes to act as an intermediary must offer contracts which maximize the expected utility of the borrowers and assure the remaining lenders of at least the reservation level of utility (r), which is determined by the riskless rate of return on the alternative project. Otherwise, agents would trade directly or another intermediary would offer an alternative contract (i.e., there is free entry into intermediation) with terms that are preferable to the n borrowers and/or the remaining $m-1$ lenders. Let $(R(\cdot), S)$ denote aspects of the two-sided contract which pertain to the borrower-intermediary relationship and $(R^*(\cdot), S^*)$ denote aspects of the two-sided contract which pertain to the intermediary-lender relationship.[10] The intermediary's problem clearly embodies optimization by all agents in the economy.

We next derive random variables which describe the income from the intermediary's portfolio. Recall that $R_i(x)$ denotes the payoff by borrower i to the intermediary if output x is realized, X_i is the random variable which describes the output x of a particular borrower i in state x, and $X_i = Y_i + Z$ from equation (1), where the Y_i are independent random variables but the

[10] $(R(\cdot), S)$ is also used in the direct investment problem in Section 3.1. We do not introduce additional notation in this Section because the structure of the problem is the same regardless of whether borrowers report to the lenders or to the bank.

X_i are not independent for $Z \neq 0$. The intermediary's income from borrower i, given transfer $R(\cdot)$, can now be defined by

$$G_i(R(\cdot);\omega) = R(X_i(\omega)), \tag{3}$$

where $\omega \in \Omega$ denotes the state of nature. Because the X_i are not independent, it follows that in general the random variables G_i are *not* independent. If the intermediary contracts with $i = 1, 2, \ldots, n$ borrowers,[11] its average income per borrower under payment schedule $R(\cdot)$ is: $G^n(R(\cdot);\omega) = \frac{1}{n}\sum_{i=1}^{n} G_i(R(\cdot);\omega)$. Denote the distribution function of $G^n(\cdot)$ by $F^n(\cdot)$.

The two-sided contract between the intermediary and each borrower, and the intermediary and the lenders, can now be defined.

Definition 3. A two-sided contract is a four-tuple $((R(\cdot), S), (R^*(\cdot), S^*))$ with the following properties:

(i) $R(\cdot)$ is an integrable positive payment function from a borrower to the intermediary such that $R(x) \leq x$ for every $x \in \mathbb{R}_+$, and S is an open subset of $\mathbb{R}_+$ which determines the set of all realizations of a borrower's project where the intermediary must monitor;

(ii) $R^*(\cdot)$ is an integrable positive payment function from the intermediary to the lenders such that $R^*(x) \leq x$. For every realization x of $G^n(\cdot)$, the payment to an individual lender is given by $\frac{n}{m-1}R^*(x)$;[12] and S^* is an open subset of $\mathbb{R}_+$ which determines the set of all realizations of the intermediary's income from the borrowers the lenders must verify.

We now derive the set of all incentive-compatible two-sided contracts. Each borrower will announce an output which minimizes its payment obligations to the intermediary. Let $\bar{x} = \arg\min_{x \in S} R(x)$ be the output that minimizes this payoff over all non-monitoring states, and recall that x is observed directly in the monitoring states S. Consequently, the announcement by a borrower is given by $\arg\min_{\hat{x} \in \{x, \bar{x}\}} R(\hat{x})$. A similar condition holds for the intermediary-lender portion of the contract (i.e., $R^*(\cdot), \bar{R}^*$). As in

[11] Note that n need not equal $\bar{n}$. In fact, this paper shows that it in general it will not be optimal for a bank to become as large as possible.

[12] $R^*(\cdot)$ is the total payment by the intermediary to lenders per borrower. Since the intermediary has a positive initial endowment, $m-1$ lenders are sufficient to finance the m projects. Thus, to derive the payment to an individual lender, multiply this amount with $\frac{n}{m-1}$.

the one-sided problem, the following condition ensures that all contracts are incentive-compatible. There exist $\bar{R}$, $\bar{R}^* \in \mathbb{R}_+$ such that $S = \{x: R(x) < \bar{R}\}$ and $S^* = \{x: R^*(x) < \bar{R}^*\}$. The set of all incentive-compatible two-sided contracts is fully specified by the four-tuple $(R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)$.

A two-sided simple debt contract is then defined as follows:

Definition 4. A contract $(R(\cdot), \bar{R})$, $(R^*(\cdot), \bar{R}^*)$ is a two-sided simple debt contract if:
(i) $R(x) = x$ for $x \in S = \{x < \bar{R}\}$ and $R(x) = \bar{R}$ if $x \in S^c = \{x \geq \bar{R}\}$; and
(ii) $R^*(x) = x$ for $x \in S^* = \{x < \bar{R}^*\}$ and $R^*(x) = \bar{R}^*$ if $x \in S^{*c} = \{x \geq \bar{R}^*\}$.
We will often denote two-sided simple debt contracts by $(\bar{R}, \bar{R}^*)$.

The intermediary's two-sided optimization problem can now be stated. Let c denote the intermediary's cost of monitoring the borrowers, and let c_n^* denote the lenders' cost of monitoring the intermediary. In Section 4 these monitoring costs shall be discussed in more in detail.

Problem 3.2. Choose incentive-compatible contracts $(R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)$ to:

$$\max \int_0^T [x - R(x)]\, dF(x)$$

subject to:

$$\frac{n}{m-1} \int_0^T R^*(x)\, dF^n(R(\cdot), \bar{R})(x) - \int_{S^*} c_n^*\, dF^n(R(\cdot), \bar{R})(x) \geq r \qquad (4)$$

$$n \left[\int_0^T [x - R^*(x)]\, dF^n(R(\cdot), \bar{R})(x) - \int_S c\, dF(x) \right] \geq r. \qquad (5)$$

Problem 3.2 states that the intermediary maximizes the expected utility of each ex-ante identical borrower subject to two constraints. (4) states that the expected payoff to the $m-1$ remaining lenders (i.e., those who did not become intermediaries) must be at least r, the level of utility available from the alternative project. (5) states that the profit from intermediation (i.e., net payoffs from the borrowers less the payoff to the lenders) must also be at least r. Note that the bank's decision variables are the loan contract $R(\cdot)$, the deposit contract $R^*(\cdot)$, and the number of projects n. The number of lenders is determined by the choice of n.

4 Optimal Investment Arrangements

The structure of the optimal investment arrangement will depend crucially on the nature of the monitoring costs, c and c_n^*, because default risk is non-trivial and monitoring will occur with positive probability. We now proceed to prove Theorem 1 which establishes that delegated monitoring is optimal when the lenders' costs of monitoring the intermediary are bounded and the variance of the nondiversifiable macroeconomic risk is sufficiently small. The proof of Theorem 1 depends on continuity of the constraints of Problem 3.2 in the face values $\bar{R}$ and $\bar{R}^*$ of the two sided debt contract, which follows from Lemma 1 in Krasa and Villamil (1991a). The strategy of the proof of the Theorem is as follows. Let $\bar{R}$ be the simple debt contract which is optimal among all one-sided schemes described by Theorem GHW in Section 3.1. We show: (i) there exists an alternative two-sided debt contract $(\bar{R}, \bar{R}^*)$ such that (4) is satisfied and binding; (ii) (5) is fulfilled but does not bind under $(\bar{R}, \bar{R}^*)$; and (iii) by increasing the face value of the lenders' debt (say to $\bar{B}^* > \bar{R}^*$) the payoff to the lenders increases.[13] Then by continuity of the constraints in the face value of the lenders' debt, a two-sided contract $(\bar{R}, \bar{B}^*)$ can be found such that both constraints are slack. Finally, by continuity of the constraints in $\bar{R}$, the face value of the borrowers' debt, $\bar{R}$, can be lowered, with both constraints still satisfied. Thus, the delegated monitor offers better contracts to agents than the best feasible direct investment contract, which proves the Theorem. The argument requires n, the number of borrowers, to be sufficiently large; a more precise characterization of n shall be provided in Section 5.

Theorem 1. *Assume that c_n^* is bounded and that the variance of Z is sufficiently small. Then delegated monitoring with two-sided debt contracts dominates direct investment.*

Proof. Consider first the investors' cost of monitoring the bank. Recall that the bank faces two types of risk: a diversifiable, project-specific risk Y_i, and a non-diversifiable macroeconomic risk Z. Thus, the bank's default probability will in general not converge to zero (even if it contracts with a

[13] In general, the lenders' payoff does not increase monotonically with $\bar{R}^*$ because the probability that lenders must verify the intermediary is an increasing function of $\bar{R}^*$. This is also true for one-sided schemes (cf., Gale and Hellwig (1985, p. 662)).

large number of borrowers). By Lemma 1 of the Appendix, the average payoff from borrowers to the intermediary converges to the expected conditional return $E[R(X_1) \mid Z]$:[14]

$$\frac{1}{n}\sum_{i=1}^{n} R(Y_i + Z) \to E[R(Y_1 + Z) \mid Z], \tag{6}$$

as $n \to \infty$. Further, by Lemma 2 in the Appendix, the probability that the return from borrowers is less than the lenders' fixed payment converges to the probability that the expected return $E[R(X_1) \mid Z]$ is less than the face value of the lenders' debt:

$$P\left(\left\{\frac{1}{n}\sum_{i=1}^{n} R(Y_i + Z) < \bar{R}^*\right\}\right) \to P\left(\left\{E[R(Y_1 + Z) \mid Z] < \bar{R}^*)\right\}\right). \tag{7}$$

Now choose $\bar{z}(\bar{R}^*)$ such that $E[R(Y_1 + \bar{z}(\bar{R}^*))] = \bar{R}^*$. Note that $\bar{z}(\bar{R}^*)$ is independent of the distribution of Z. Furthermore,

$$E\left[R(Y_1 + \bar{z}(\bar{R}^*))\right] = E\left[R(Y_1 + Z) \mid Z = \bar{z}(\bar{R}^*)\right].^{15}$$

Note that $\bar{z}(\cdot)$ is a "cutoff value" in the distribution of Z which separates solvency states from insolvency states in the limit. Since $R(\cdot)$ is monotonic the right-hand-side of (7) equals $P\left(\left\{Z \leq \bar{z}(\bar{R}^*)\right\}\right)$, which we shall show is the *bank's default probability* in the limit.

Next, consider the bank's payoff to a lender when its portfolio is large. From (6) and from continuity of $R^*(\cdot)$ it follows that

$$R^*\left(\frac{1}{n}\sum_{i=1}^{n} R(Y_i + Z)\right) \to R^*\left(E[R(Y_1 + Z) \mid Z]\right)$$

as $n \to \infty$. Lebesgue's dominated convergence Theorem therefore implies

$$\lim_{n\to\infty} \int R^*\left(\frac{1}{n}\sum_{i=1}^{n} R(Y_i(\omega) + Z(\omega))\right) dP(\omega)$$
$$= \int R^*\left(E[R(Y_1 + Z) \mid Z](\omega)\right) dP(\omega) \tag{8}$$

[14]For all random variables X, Y denote by $E[X \mid Y]$ the conditional expectation of X with respect to Y (which is a random variable, measurable with respect to the information contained in Y). In particular, let $\omega \in \Omega$ be an elementary event for which $Y(\omega) = y$. Then $E[X \mid Y = y] = E[X \mid Y](\omega)$. If X and Y have only a countable number of different values then this corresponds to the elementary definition of a conditional expectation.

[15]This follows from our strong independence assumption. See footnote 5 and the proof of Lemma 1.

Substituting the distribution of $\frac{1}{n}\sum_{i=1}^{n} R(Y_i(\omega)+Z(\omega))$ and the distribution of Z for P in (8) yields

$$\lim_{n\to\infty}\int R^*(x)\,dF^n(R(\cdot))(x) = \int R^*\left(E[R(Y_1+Z)\mid Z=z]\right)\,dH(z) \qquad (9)$$

where H denotes the distribution of Z. For example, if there is *no* macroeconomic risk (i.e., $Z=0$), the right-hand-side of (9) is given by $R^*\left(E[R(X_1)]\right)$ so when $\bar{R}^* < E[R(X_1)]$ the expected return in the limit is given by $\bar{R}^*$ and the lenders receive the face value of the debt with certainty.

A two-sided contract which dominates the one-sided, direct investment contract can now be constructed. Let $\varepsilon > 0$ be some arbitrary constant. First, choose $\bar{B}^*$ such that $r < \bar{B}^* < E[R(X_1)]$. Then (7) implies that the bank's default probability is less than ε for large n, if $P(\{Z \le \bar{z}(\bar{B}^*)\}) < \varepsilon$, i.e., the probability that the realization is in the "tail" of the distribution of Z is sufficiently small. Note, that $\bar{z}(\bar{B}^*) < 0$.[16] Thus, there exists a $\delta > 0$ such that whenever $\mathrm{var}(Z) < \delta$ we get $P(\{Z \le \bar{z}(\bar{B}^*)\}) < \varepsilon$.[17] Thus, the lender's expected costs of monitoring the bank are bounded above by εc_n^* for large n. For similar reasons, the lenders' payoff is bounded from below by $\bar{B}^*(1-\varepsilon)$ for sufficiently large n.

If ε is sufficiently small, (7) and (9) imply that constraint (4) is fulfilled, but does not bind for the two-sided contract $(\bar{R}, \bar{B}^*)$. By continuity of (4) with respect to $\bar{B}^*$ (see Krasa and Villamil (1991, Lemma 1)), there exists a face value $\bar{R}^* < \bar{B}^*$ such that (4) binds for the two-sided contract $(\bar{R}, \bar{R}^*)$. We next show that (5) is fulfilled under contract $(\bar{R}, \bar{R}^*)$, but does not bind. Recall that the bank's default probability is less than ε. Thus, $\int_{S^*} c_n^*\,dF^n < \varepsilon c_n^*$ for the contract with face value $\bar{B}^*$. Since $\bar{R}^* < \bar{B}^*$, the bank's default probability is lower with $\bar{R}^*$. Thus, by choosing ε sufficiently small and since c_n^* is bounded we can ensure that $\int_S c\,dF > \int_{S^*} c_n^*\,dF^n$ for all sufficiently large n.[18] This and the fact that (4) binds implies

$$n\int_0^T R^*(x)\,dF^n < (m-1)(r+\int_S c\,dF). \qquad (10)$$

[16]We normalize the mean of the macroeconomic shock Z to zero, so $\bar{z}(\bar{B}^*) \le 0$ denotes a recession.

[17]This is possible since $z(\bar{B}^*)$ is independent of the distribution of Z.

[18]The inequality indicates that the intermediary's expected cost of monitoring the borrowers is higher than the lenders' expected cost of monitoring the intermediary.

Consequently,

$$n\left[\int_0^T [x - R^*(x)]\, dF^n - \int_S c\, dF\right]$$
$$> nE[R(X_1)] - m\int_S c\, dF - (n-1)r \geq mr - (m-1)r = r.$$

The first inequality follows from (10) and from the fact that $\int x\, dF^n = E[G^n] = E[R(X_1)]$, which is the expected value of an aggregate version of equation (3). The second inequality follows because $\bar{R}$ must fulfill (2) by assumption. Now increase $\bar{R}^*$ slightly. Then by the continuity of the constraint and by the construction of $\bar{R}^*$ the lenders' payoff increases and thus both constraints can be made slack. This proves Theorem 1 because there exists some surplus that can be redistributed to borrowers by lowering the face value of their debt $\bar{R}$.

Theorem 1 establishes optimality of delegated monitoring schemes if n is sufficiently large and if the variance of the macroeconomic shock is sufficiently small. However, it does not follow that it is optimal for the bank to be as large as possible. Indeed, Theorem 1 suggests that an optimal bank size may exist because as the bank increases its portfolio size there are gains from default risk reduction but losses from increased monitoring costs. In Theorem 2 we characterize these gains and losses more precisely. However, before doing so we first relate Theorem 1 to the previous literature on delegated monitoring. Specifically, we focus on the bank size and industry structure predictions implicit in previous models.

Diamond (1984) and Williamson (1986) use a law of large numbers argument to prove the optimality of delegated monitoring (i.e., financial intermediation) in an economy with bounded costs and *no* macroeconomic risk. Because the probability that a bank fails is zero in the limit in their models, the lenders' expected costs of monitoring the bank are zero. Clearly, a bank that operates in such an environment can always reduce the expected monitoring costs borne by lenders by increasing its size. Thus, "big banks are always better," and the model predicts banks of large but indeterminate size.[19] This size prediction is implicit in the Diamond and Williamson models, and stems from the fact that increasing returns to scale are inherent in

[19]Because the set of borrowers is infinite, it is possible to get multiple banks that are indeterminately large in this framework. However, the argument requires that the infinite

the framework they consider. Specifically, in their models a bank can always both decrease the riskiness of its portfolio *and* reduce monitoring costs by contracting with additional borrowers. An obvious question, therefore, is: Does delegated monitoring in an economy with non-diversifiable portfolio risk also give rise to increasing returns to scale in intermediation, and hence indefinitely large banks? The answer to this question depends on the specification of monitoring costs.

Consider first a "best case" situation where lenders face a fixed cost of monitoring a bank in default states. Specifically, let $c_n^* = k$, where k is a positive constant which is independent of the size of the bank. In this case, (7) from Theorem 1 implies that the bank's default probability converges to $P(\{Z \leq \bar{z}\})$. This follows from the fact that a bank's default probability decreases (in general) as it contracts with more borrowers because idiosyncratic risk is diversified away. The non-diversifiable macroeconomic risk obviously remains.[20] To make this argument more precise, let $p_{\hat{n}}$ denote the bank's default probability when it has a portfolio of size $\hat{n}$, where $\hat{n} < n$. The lenders' expected costs of monitoring the bank under this cost structure are $p_{\hat{n}} c_n^* = p_{\hat{n}} k$. Since the bank's default probability when its portfolio size is $\hat{n}$ is at least as great as its default probability in the limit (i.e., $p_{\hat{n}} \geq \lim_{n \to \infty} p_n$), the lenders' expected costs of monitoring the intermediary are lower for larger banks. It follows from this observation that the delegated monitoring model with non-diversifiable portfolio risk *and* constant monitoring costs will generally also display increasing returns to scale in intermediation (because increasing the bank's portfolio size does not raise monitoring costs but it may lower the bank's default probability). Consequently, like Diamond and Williamson, the optimal bank size under constant monitoring costs is indeterminately large.

Now consider a polar opposite "worst case" situation, where the lenders'

---

set of borrowers be partitioned into an infinite number of subsets where each infinite subset of borrowers contracts with a particular delegated monitor. This argument does not appear to be a plausible explanation of the observed co-existence of multiple banks. Of course, the models were not designed to explain this observation.

[20]Convergence in equation (7) need not be monotonic. Thus, there may exist points of non-monotonic convergence where even under constant monitoring costs the optimal bank size is finite if the macroeconomic risk is non-trivial. We therefore assume without loss of generality that the bank's probability of default is always bounded from below by $P(\{Z \leq \bar{z}\})$, which is the default probability of a bank of infinite size with macroeconomic risk but no idiosyncratic risk.

monitoring costs are unbounded. Krasa and Villamil (1991, Theorem 1) analyze this problem when there is no macroeconomic risk. They show that even if costs are unbounded but do not increase at an exponential rate, delegated monitoring with two-sided simple debt contracts still dominates direct investment. However, the non-trivial macroeconomic risk which we consider in this paper complicates this problem considerably. Specifically, in the limit the lenders' expected monitoring costs are given by $\lim_{n\to\infty} p_n c_n^*$. Clearly, if c_n^* is unbounded this product converges to infinity when the bank's portfolio is subject to non-diversifiable macroeconomic risk (because $\lim_{n\to\infty} p_n > 0$). Thus, constraint (4) from Problem 3.2 is violated for sufficiently large n, and this implies that delegated monitoring is not feasible with unbounded costs, non-diversifiable macroeconomic risk, and a sufficiently large portfolio size. This argument suggests that an optimal portfolio (or bank) size may exist because the feasibility of delegated monitoring depends on n.

Consider now the problem of whether or not a bank of a given size ($\hat{n}$) should contract with additional borrowers, thus increasing its scale. Suppose that monitoring costs are bounded but not constant. Theorem 1 establishes that delegated monitoring is optimal if the variance of the macroeconomic risk is sufficiently small, but this does *not* imply that the bank should be as large as possible. When there is non-trivial macroeconomic risk and the bank has a portfolio of size $\hat{n}$, the bank must consider two factors when deciding whether or not to increase its scale.

(i) Adding additional projects to its portfolio decreases the bank's default risk. This decrease is given by the difference between the bank's probability of default in the limit (i.e., $\lim_{n\to\infty} p_n$) and its probability of default at portfolio size $\hat{n}$ (i.e., $p_{\hat{n}}$); but for $\hat{n}$ sufficiently large, the gains from reducing default risk by adding additional projects are essentially zero.

(ii) Adding additional projects to the banks' portfolio raises the lenders' costs of monitoring the bank.

Thus, the crucial question is: For what cost structures do the gains from reduced default risk dominate the losses from increased monitoring costs when additional projects are added?

5 Optimal Intermediary Size

We now obtain the main results of the paper: analytic predictions for both the optimal size of a financial intermediary *and* for the size distribution of banks in a Pareto efficent industry. In order to analyze the problem in detail (and answer the question posed above), we must provide a precise quantitative characterization of the gains from diversification as the size of the bank increases. To this end we construct a rate function which measures the speed at which the bank's idiosyncratic portfolio risk is eliminated when the size of its portfolio is increased. The *Theory of Large Deviations* (cf., Varadhan (1984)) provides the formal structure.

Consider first the bank's portfolio diversification problem for the case of a fixed realization z of Z. The argument will be generalized to permit any z by integrating over the distribution of all possible realizations of z. Then $X_i^z = Y_i + z$ are independent random variables (given that z is fixed) since the Y_i are independent. Since the random variables X_i^z are independent, the law of large numbers holds. The *large deviation principle* gives a rate function (cf., Varadhan (1984), Theorem 3.1) which provides a measure of the speed of convergence in the law of large numbers. The rate function implies that for every $z < E_z[R(X_i^z)]$, the probability that a realization is in the tail of the distribution converges exponentially to zero. Formally

$$P\left(\left\{\frac{1}{n}\sum_{i=1}^{n} X_i^z \leq a\right\}\right) \leq e^{-\mathcal{I}_z(a)n}, \tag{11}$$

where $\mathcal{I}_z(a) > 0$ is the *rate function* which gives the speed of convergence of the distribution (i.e., a measure of how rapidly contracting with additional borrowers reduces the bank's default risk).

The rate function $\mathcal{I}_z(\cdot)$ is derived from the moment-generating function of a random variable. Let $M_z(\theta)$ denote the *moment-generating function* of the distribution of $R(X_i^z)$, where μ_z denotes the distribution. Then

$$M_z(\theta) = \int e^{\theta x}\, d\mu_z(x).^{21}$$

The rate function is found by solving the following maximization problem:

$$\mathcal{I}(a) = \max_{\theta \in \boldsymbol{R}} \theta a - \log M_z(\theta). \tag{12}$$

[21] $M_z(\theta)$ is called the moment generating function since the kth derivative of $M_z(\theta)$ evaluated at $\theta = 0$ gives exactly the kth moment of μ.

We now show that $\mathcal{I}(a) > 0$ for every $a < E(X_i^z)$. For fixed a let $f(a,\theta) = \theta a - \log M_z(\theta)$. Since $f(a,0) = 0$, it is sufficient to show that $\frac{\partial}{\partial a} f(a,0) \neq 0$, which can be easily verified:

$$\frac{\partial}{\partial a} f(a,\theta) = a - \frac{\int x e^{\theta x}\, d\mu_z(x)}{\int e^{\theta x}\, d\mu_z(x)}.$$

Consequently, $\frac{\partial}{\partial a} f(a,0) = a - EX_i^z \neq 0$.[22]

We have shown that the probability $P\left(\left\{\frac{1}{n}\sum_{i=1}^n R(Y_i + z) \leq \bar{R}^*\right\}\right)$ converges exponentially to $E[R(Y_i + z)]$ for fixed z such that $z > \bar{z}(\bar{R}^*)$ (i.e., if the macroeconomic shock is not too severe). We now make the convergence argument independent of z. The primary technical problem is that the behaviour of the rate function must be analyzed as z comes close to $\bar{z}(\bar{R}^*)$. Recall that for $z = \bar{z}(\bar{R}^*)$, the face value of the lenders' contract, $\bar{R}^*$, is exactly the expected value of $R(Y_i + z)$. Clearly, the probability of observing realizations less than or equal to the expected value of independent random variables does not converge to zero. However, in Lemma 3 we show that for $z \geq \bar{z}(\bar{R}^*)$, and z sufficiently close to $\bar{z}$, the rate function $\mathcal{I}_z(\bar{R}^*)$ is bounded from below by $k(z - \bar{z})^2$, where $k > 0$. In particular, from (A.15) and (A.16) in the Appendix it follows that we can choose for k any number smaller than $\frac{1}{2 \operatorname{var} R(Y_i + \bar{z})}$.[23] Thus, the smaller the variance of the idiosyncratic risk the faster convergence. Thus, the rate function converges to zero at the speed $(z - \bar{z})^2$ for $z \to \bar{z}$. Integrating over z, the probability of default by a bank of size n, given that $Z \geq \bar{z}$, is:

$$\int_{\bar{z}}^{\infty} e^{-\mathcal{I}_z(\bar{R}^*)n}\, dH(z). \tag{13}$$

An important technical result, which is essential for proving that an optimal bank size exists (Theorem 2), can now be stated.

Proposition 1. *There exist constants $k_i > 0$, $i = 1, 2$ such that the probability of default by a bank of size n, given that $Z \geq \bar{z}$, is bounded from above by $\frac{k_1}{\sqrt{n}} + e^{-nk_2}$.*

[22] This follows from the fact that $\int x e^{\theta x}\, d\mu_z(x)$ evaluated at $\theta = 0$ is the expected value of X_i^z, since μ_z is the distribution of X_i^z. Furthermore $\int e^{\theta x}\, d\mu_z(x)$ evaluated at $\theta = 0$ is one.

[23] Apply Lemma 3 as in Proposition 1 and choose $X_a = R(Y_i + z)$.

The proof of Proposition 1, which provides a speed of portfolio convergence result, is in the Appendix. Note that in the bound in Proposition 1, the term e^{-nk_2} converges to zero much faster than $\frac{k_1}{\sqrt{n}}$. By the law of large numbers, a bank of infinite size will never default if $Z \geq \bar{z}$. Thus, the bound implies that a bank of size n can lower its default probability by only approximately $\frac{k_1}{\sqrt{n}}$ if it becomes infinitely large (given that $Z \geq \bar{z}$) because the second term is approximately zero for large n. Consequently, $\frac{k_1}{\sqrt{n}}$ is our desired measure of the gain from default risk reduction arising from additional diversification.

The main result of the paper can now be stated.

Theorem 2. *Let c_n^* denote the cost of monitoring a bank of size n. Let $c_\infty^* = \lim_{n\to\infty} c_n^*$,[24] and assume that $c_\infty^* - c_n^*$ converges to zero at a slower rate than $\frac{1}{\sqrt{n}}$. Then it is never optimal for a bank to become infinitely large. Thus, there exists an optimal size for the bank.*

Proof. Assume by way of contradiction that it is optimal for the bank to become infinitely large. By Proposition 1, the bank's probability of default is bounded above by

$$p_n = \frac{k_1}{\sqrt{n}} + e^{-k_2 n} + P(\{Z \leq \bar{z}\}).^{25} \tag{14}$$

By Lemma 2, the bank's default probability in the limit is given by $P(\{Z \leq \bar{z}\})$. Let n be arbitrary. Now compare the expected monitoring costs for a bank of size n with those of a bank of infinite size. By (14), the lenders' expected costs of monitoring a bank of size n (i.e., $p_n c_n^*$) are at least

$$\left[\frac{k_1}{\sqrt{n}} + e^{-k_2 n}\right] c_n^* + P(\{Z \leq \bar{z}\}) c_n^*. \tag{15}$$

The expected costs of monitoring a bank of infinite size are at most

$$P(\{Z \leq \bar{z}\}) c_\infty^*. \tag{16}$$

[24] If c_n^* is unbounded then $c_\infty^* = \infty$, and clearly Theorem 2 holds.

[25] The first two terms are the bound given by Proposition 1 for all states where $Z \geq \bar{z}$. Assuming that the probability of default in all states $Z \leq \bar{z}$ is one (which is clearly an upper bound for these states), we get the third term.

If it were optimal for the bank to become infinitely large, then at least for large n the expected costs of monitoring the bank per lender must decrease if the size of the bank is increased from n to infinity. By (15) and (16) the expected monitoring costs will decrease if[26]

$$\left[\frac{k_1}{\sqrt{n}} + e^{-k_2 n}\right] c_n^* > P(\{Z \leq \bar{z}\})\,(c_\infty^* - c_n^*)\,. \tag{17}$$

By the assumption of the Theorem, $c_\infty^* - c_n^*$ converges to zero at a slower rate than $\frac{1}{\sqrt{n}}$. Thus, for every $M > 0$ there must exist an $\bar{n}$ such that $c_\infty^* - c_n^* \geq \frac{M}{\sqrt{n}}$ for all $n \geq \bar{n}$. This, and equation (17) yields

$$\left[\frac{k_1}{\sqrt{n}} + e^{-k_2 n}\right] c_n^* > P(\{Z \leq \bar{z}\})\frac{M}{\sqrt{n}}. \tag{18}$$

Since M can be chosen arbitrarily, there exist values such that inequality (18) is violated.[27] The bank cannot be infinitely large, and the Theorem is thus proved.

Theorem 2 establishes that when the lenders' monitoring costs depend on the bank's portfolio size, an optimal determinate bank size can be computed, under the assumption that the rate of increase of the lenders' monitoring costs converges to zero at a sufficiently slow rate. This assumption is fairly weak. In particular, it is fulfilled if the rate of increase of c_n^* is of the order $\frac{1}{n^{1.5}}$.[28] Most economically plausible cost structures will satisfy this assumption. An alternative way to interpret Theorem 2 is that the case of constant lender monitoring costs (on which others have relied) is rather special. In particular, we argued in Section 4 that under constant monitoring costs a bank's size is indeterminate (because of increasing returns to scale in delegated monitoring). However, with a slightly changed (i.e., size dependent) cost structure an optimal bank size exists. Thus, the indeterminacy (and hence increasing returns to scale for all n) result does not seem to be very robust. Theorem 2

[26]This follows from (16) − (15) < 0.

[27]Multiply both sides of (18) by $\sqrt{n}$ to obtain $\left[k_1 + \sqrt{n}e^{-k_2 n}\right] c_n^* > P(\{Z \leq \bar{z}\})M$. Let $n \to \infty$. Then, $k_1 c_\infty^* \geq P(\{Z \leq \bar{z}\})M$, which cannot hold for every M. Thus (18) must be violated for all sufficiently large n.

[28]Note that $c_\infty^* - c_n^*$ is approximately $\frac{1}{\sqrt{n}}$. Differentiating with respect to n (ignoring that n is an integer) of course yields $\frac{1}{n^{1.5}}$.

also shows that even when the bank's portfolio is subject to non-diversifiable macroeconomic risk (as long as the variance of the risk is not too large so intermediation remains optimal), a bank of size n can only improve upon its default probability by at most $\frac{k_1}{\sqrt{n}}$. Increasing bank size after some critical $\bar{n}$ is not optimal because it leads to increased monitoring costs, but there are very limited gains from default risk reduction. Thus, even very "small" banks (given Z) may improve upon direct investment when intermediation is optimal.

6 Testable Implications of the Theory

Theorem 2 has the following testable implications. First, bank size is determinate and inversely related to the bank's exposure to macroeconomic risk. In a large economy like the U.S. where aggregate macroeconomic shocks have different effects on different regions of the country, our model predicts that banks of different sizes will coexist across locales. In particular, if different regions of the country have different effective macroeconomic shocks (i.e., Z's), the model predicts that across regions both large "money centre banks" that are very well diversified and smaller "local banks" that are less well diversified will co-exist. "Money centre banks" may have been able to lower their exposure to macroeconomic risk, by evading portfolio restrictions via holding companies or because they operate in regions of the country with better diversified economic bases. Our model predicts that these better diversified (i.e., low Z) banks will be larger than "local banks" (with higher Z's). This follows from (18) in the proof of Theorem 2 because $P(\{Z \leq z\})$ is lower for a better diversified bank, since the bank's exposure to macroeconomic shocks is effectively less severe. Therefore, the bank size (n) that violates (18) is necessarily higher. Williamson (1989) provides an interesting discussion of stylized facts regarding the structure of U.S. versus Canadian banks. Historically, Canadian banks have had fewer portfolio (e.g., branching restrictions) and hence lower Z's than U.S. banks. As our theory predicts, there have been fewer banks in Canada of larger size (adjusted for population differences).

The second testable implication of our theory pertains to industry structure, i.e., limits on the number and size distribution of firms that are present in equilibrium. Theorem 1 establishes that intermediation (banking) improves social welfare. In other words, there is a some level of intermediation

services in an economy that is socially optimal (given preferences, costs, probability distributions, and alternative opportunities). In contrast, Theorem 2 proves that there is a specific bank size that is optimal. Although the notions of bank size and industry structure are closely related, they need not be identical. For example, suppose the socially optimal "industry" (i.e., economy) level of welfare improving intermediation services is twenty units of input capital and the optimal size for all banks is to provide two units of capital. Clearly, the optimal industry structure is then ten banks. What causes some banks to be of similar size in our model? The result that within a region (or among banks with similar effective portfolio restrictions) banks with similar "local" idiosyncratic risk characteristics will have a common size (given their Z) follows immediately from equation (18), because k_1, M and $P(\{Z \leq \bar{z}\})$ will be similar for such banks. Why might many moderately sized local banks coexist in the same region? This again follows from Theorem 2 because the Theorem establishes that after some critical size—increasing a bank's scale of operation further is not profit maximizing. Finally, why might we expect to observe many small "local banks" and fewer large "money centre banks?" The high frequency of smaller local banks (relative to large money centre banks) stems from the fact that banks with higher Z's cannot achieve sufficient portfolio diversification in their locale to justify the additional monitoring costs associated with increasing their scale of operation. Multiple banks operating at the efficient scale within a particular locale provide welfare improving intermediation services optimally (given the risk and cost structure in the economy). Differences in bank size and distribution across locales stem entirely from different effective Z's in (18) (i.e., different exposure to macroeconomic risk in local and money centre banks).

7 Concluding Remarks

In this paper we develop a theory of bank size distribution with testable implications. We regard the theoretical model to be useful for two reasons. First, there is a longstanding debate in the banking literature about the market structure of the financial industry. Some have argued that the banking industry is inherently non-competitive and hence must be regulated to protect consumers from the evils of banks' market power. In contrast, others have argued that the banking system is fragile and hence must be protected

from the destabilizing forces of competition. Because we derive optimal financial contracts from Pareto problems, the allocations that we obtain are necessarily Pareto efficient. An interesting problem for future research is to compare the structure of a banking industry where firms have market power with the Pareto efficient structure implied by Theorem 2 (given parametric specifications for preferences, costs, probability distributions, and alternative assets).

Second, the model may prove useful in understanding changes in the European banking system that will undoubtedly result from the EMU. In the previous Section we discussed bank size and industry structure predictions for the U.S. economy. For example, our model predicts that largely agricultural sections of the U.S. like the Mid-west will have many moderate or small sized banks because this region is subject to marcoeconomic shocks that are difficult to diversify (especially given portfolio restrictions imposed by bank regulators). However, the model predicts money centre banks that operate in more economically diversified regions of the country (and that have been better able to evade portfolio restrictions) will be larger. In 1999 when the EMU begins, twelve countries will form a "federal Europe." What will be that nature of the EMU banking regulations imposed by the central bank or by the governments of the individual countries? Are the banks of some countries (specifically, those with better diversified economies) destined to become "money centre" banks while the banks of other (less well diversified) economies are destined to become small "local" banks? This question is important because—although our model permits existence of both multiple banks of the same size, and perhaps more interestingly the co-existence of banks of different sizes—larger (better diversified banks) have a lower default probability than smaller banks. Are members of the EMU with less well diversified economies, that might wish to encourage their banks to lend domestically for development reasons, destined to become problematic members of the Union from the outset?

8 Appendix

We first derive a "law of large numbers" for random variables with correlation. The argument works essentially as follows: For any fixed realization z of Z we can apply the law of large numbers to the random variables $Y_i + z$

and show that $\frac{1}{n}\sum_{i=1}^{n} Y_i + z$ converges to z except for a set N_z which has measure zero with respect to the probability P_1 (recall that P is the product of the probabilities P_1 and P_2). We then use Fubini's Theorem (i.e., integrate with respect to P_2) to show that $\bigcup_{z\in\mathbb{R}} N_z$ has measure zero with respect to the original probability P. We now state Lemma 1.

Lemma 1. *Let Y_i, $i \in \mathbb{N}$ and Z be independent random variables. Assume that the Y_i are identically distributed. Let $X_i = Y_i + Z$, for every i and let $R(\cdot)$ be a simple debt contract. Then $\lim_{n\to\infty} \sum_{i=1}^{n} R(X_i)(\omega) = E\left[R(Y_1) \mid Z\right](\omega)$ for almost every $\omega \in \Omega$.*

Proof. Let z be a realization of Z. Let $E[R(X_1) \mid Z]$ be the conditional probability of X_1 with respect to Z. Let $\omega_2 \in \Omega_2$. Define the probability space $(\Omega_{\omega_2}, \mathcal{A}_{\omega_2}, P_{\omega_2})$ as follows. Recall that $\Omega = \Omega_1 \times \Omega_2$. Then let $\Omega_{\omega_2} = \Omega_1 \times \{\omega_2\}$. Let $\mathcal{A}_{\omega_2}$ be the set of all events $A \times \{\omega_2\}$, where A is an event in Ω_1. Finally, let $P_{\omega_2}(A \times \{\omega_2\}) = P_1(A)$. Let $z = Z(\omega_2)$. Then $R(X_i) = R(Y_i + z)$ are independent random variables on $(\Omega_z, \mathcal{A}_z, P_z)$. We can therefore the apply the law of large numbers to the X_i^z, $i \in \mathbb{N}$. Thus,

$$\lim_{n\to\infty} \sum_{i=1}^{n} R(X_i(\omega)) = \int_{\Omega_{\omega_2}} R(X_1(\omega) + z)\, dP_{\omega_2}(\omega) \qquad (A.1)$$

for all $\omega \in \Omega_{\omega_2}$ except for a set $N_{\omega_2} \in \mathcal{A}_{\omega_2}$ with $P_{\omega_2}(N_{\omega_2}) = 0$. Further,

$$\int_{\Omega_{\omega_2}} R(X_1(\omega))\, dP_{\omega_2}(\omega) = E[R(X_1) \mid Z](\omega), \qquad (A.2)$$

for almost every $\omega \in \Omega_{\omega_2}$, because Y_i and Z are independent.[29]

[29]Note that $\omega \mapsto \int_{\Omega_{\omega_2}} R(X_1(\omega))\, dP_{\omega_2}(\omega)$ is measurable with respect to Z. Thus, to show that we have a version of the conditional expectation it is sufficient to prove that

$$\int_A \int_{\Omega_{\omega_2}} R(X_1(\omega))\, dP_{\omega_2}(\omega) = \int_A R(X_1(\omega))\, dP(\omega),$$

for every event A which is measurable with respect to Z (i.e., which is of the form $Z^{-1}(B)$ where B is a measurable subset of $\mathbb{R}$). However, since Z is constant on Ω_1, all such sets are of the form $\Omega_1 \times C$, where C is a measurable subset of Ω_2. The result then follows from Fubini's Theorem since integrating over Ω_1 and then over Ω_{ω_2} (which is essentially Ω_1) is the same as integrating over Ω_1 once.

Let H be the distribution of Z. It remains to show that $N = \bigcup_{\omega_2 \in \Omega_2} N_{\omega_2}$ has measure zero. This follows from Fubini's Theorem. The set D of all ω where $\frac{1}{n}\sum_{i=1}^{n} R(X_i)$ does not converge is given by

$$D = \left\{ \omega : \liminf_n \frac{1}{n} \sum_{i=1}^{n} R(X_i) \neq f \text{ and } \limsup_n \frac{1}{n} \sum_{i=1}^{n} R(X_i) \neq f \right\},$$

where $f = E[R(X_1) \mid Z]$. Since the limsup and the liminf of a sequence of measurable functions is measurable, it follows that D is measurable. Further, $D \cap \Omega_{\omega_2} = N_{\omega_2}$ since N_{ω_2} is exactly the set of all $\omega \in \Omega_{\omega_2}$ where the sequence does not converge. Thus $D = N$ and Fubini's Theorem implies

$$P(N) = \int_{\Omega_2} P_1(N_{\omega_2})\, dP_2(\omega) = 0.$$

This concludes the proof of the Lemma.

We next prove a technical result necessary for the proof of Theorem 1 (i.e., a convergence result for the bank's probability of default).

Lemma 2. *Let X_i, Y_i and Z be as in Lemma 1. Assume that the distribution of Z is non-atomic, i.e., $P(\{Z = z\}) = 0$ for every $z \in \mathbb{R}$. Let $\bar{R}^*$ be the face value of the lenders' simple debt contract. Then*

$$P\left(\frac{1}{n} \sum_{i=1}^{n} R(Y_i + Z) \leq \bar{R}^* \right) \to P\left(E[R(Y_1 + Z) \mid Z] \leq \bar{R}^* \right).$$

Proof. Let $f_n = \frac{1}{n}\sum_{i=1}^{n} R(Y_i + Z)$, and let $f = E[R(Y_1 + Z) \mid Z]$. By (6) we get $f_n(\omega) \to f(\omega)$ for almost every ω. Thus, f_n converges to f in distribution[30] by Proposition 24.12 of Parthasarathy (1977). Let h be an indicator function of the interval $(-\infty, \bar{R}^*)$, i.e., $h(x) = 1$ for every $x \in (-\infty, \bar{R}^*)$ and $h(x) = 0$, otherwise. Note that h is only discontinuous at $\bar{R}^*$. Since the distribution of Z is continuous, the point $\{\bar{R}^*\}$ has probability zero. Corollary 1 of Billingsley (1968, p. 31) therefore implies that $h(f_n)$ converges

[30] Convergence in distribution means the following: Let F_n denote the distribution function of f_n for every $n \in \mathbb{N}$, and let F be the distribution function of f. Then f_n converges to f in distribution, if $\int g(x)\, dF_n(x) = \int g(x)\, dF(x)$ for every bounded and continuous function g on $\mathbb{R}$.

in distribution to $h(f)$. Let $H(F_n)$ denote the distribution of $h(f_n)$, let $H(F)$ denote the distribution of $h(f)$, let F_n denote the distribution of f_n, and let F be the distribution of f. Then

$$\begin{aligned}\lim_{n\to\infty} \int h(x)\, dF_n(x) &= \lim_{n\to\infty} \int x\, H(F_n)(x) \\ &= \lim_{n\to\infty} \int x\, dH(F)(x) = \lim_{n\to\infty} \int h(x)\, dF(x),\end{aligned}$$

where the second equality follows from the fact that $h(f_n)$ converges to $h(f)$ in distribution. This proves the Lemma 2.

We next prove a technical result on the convergence of the rate function $\mathcal{I}_z(\bar{R}^*)$.

Lemma 3. *Let X_a, $a \geq \bar{a} > 0$, be a collection of random variables such that a is the expected value of X_a for every $a \geq \bar{a}$. Assume that the moment generating function $M_a(\bar{a})$ is thrice continuously differentiable in a neighborhood $[\bar{a}, a)$ of $\bar{a}$. Let μ_a be the distribution of X_a. Assume that the support of μ_a is contained in the compact interval $[T_1, T_2]$ for every $a \geq \bar{a}$. Then there exists a constant $k > 0$ such that $\mathcal{I}_a(\bar{a}) \geq k(a - \bar{a})^2$ for all a which are sufficiently close to $\bar{a}$.*

Proof. Define

$$f(a, \theta) = \theta\bar{a} - \log M_a(\theta), \tag{A.3}$$

where $M_a(\theta)$ is the moment generating function $M_a(\theta) = \int e^{\theta x}\, d\mu_a(x)$. Note that $f(a, 0) = 0$. Furthermore,

$$\frac{\partial}{\partial\theta} f(a, \theta) = \bar{a} - \frac{\frac{\partial}{\partial\theta} M_a(\theta)}{M_a(\theta)}. \tag{A.4}$$

Let F_a denote the distribution function of μ_a (i.e., $F_a(t) = \mu_a((-\infty, t])$. Partial integration of $M_a(\theta)$ yields

$$\begin{aligned}M_a(\theta) = \int_{T_1}^{T_2} e^{\theta x}\, d\mu_a(x) &= e^{\theta x} F_a(x)\big|_{T_1}^{T_2} - \theta \int_{T_1}^{T_2} e^{\theta x} F_a(x)\, dx \\ &= e^{\theta T_2} - \theta \int_{T_1}^{T_2} e^{\theta x} F_a(x)\, dx \end{aligned} \tag{A.5}$$

Furthermore, by partial intergration we also get

$$\int_{T_1}^{T_2} F_a(x)\,dx = xF_a(x)|_{T_1}^{T_2} - \int_{T_1}^{T_2} x\,dF_a(x) = T_2 - a, \qquad (A.6)$$

and

$$\int_{T_1}^{T_2} xF_a(x)\,dx = \frac{1}{2}x^2F_a(x)|_{T_1}^{T_2} - \frac{1}{2}\int_{T_1}^{T_2} x^2\,dF_a(x) = \frac{1}{2}\left(T_2^2 - E(X_a^2)\right). \qquad (A.7)$$

Let $\theta^*(a)$ denote the value of θ which maximizes (A.3) for fixed a. Thus, from (A.4) and the Implicit Function Theorem we get

$$\frac{d}{da}\theta^*(a) = -\frac{\bar{a}\frac{\partial}{\partial a}M_a(\theta) - \frac{\partial^2}{\partial\theta\partial a}M_a(\theta)}{\bar{a}\frac{\partial}{\partial\theta}M_a(\theta) - \frac{\partial^2}{\partial\theta^2}M_a(\theta)}. \qquad (A.8)$$

We now evaluate $\frac{d}{da}\theta^*(a)$ at $a = \bar{a}$. Thus, since $\theta^*(\bar{a}) = 0$ by (A.4),[31] we must evaluate the right-hand side of (A.8) at $\theta = 0$ and $a = \bar{a}$. Taking the partial derivatives of $M_a(\theta)$ in (A.5), evaluating them at $\theta = 0$ and $a = \bar{a}$ and using (A.6) proves that $\frac{\partial}{\partial a}M_{\bar{a}}(0) = 0$. Furthermore,

$$\frac{\partial^2}{\partial\theta\partial a}M_{\bar{a}}(0) = -\frac{\partial}{\partial a}\int_{T_1}^{T_2} F_a(x)\,dx = 1,$$

where the last equality follows from (A.6). Thus, the numerator of (A.8) is -1. We next derive the denominator of (A.8). Note that (A.5) implies,

$$\frac{\partial}{\partial\theta}M_{\bar{a}}(0) = T_2 - \int_{T_1}^{T_2} F_{\bar{a}}(x)\,dx = T_2 - (T_2 - \bar{a}) = \bar{a},$$

where the second equality follows from (A.6); and

$$\frac{\partial^2}{\partial\theta^2}M_{\bar{a}}(0) = T_2^2 - 2\int_{T_1}^{T_2} xF_{\bar{a}}(x)\,dx = T_2^2 - (T_2^2 - E(X_a^2)) = E(X_a^2),$$

where the second equality follows from (A.7). Thus, (A.8) implies

$$\frac{d}{da}\theta^*(\bar{a}) = \frac{1}{\bar{a}^2 - E(X_{\bar{a}}^2)} = -\frac{1}{\mathrm{var}(X_a)}. \qquad (A.9)$$

[31] This follows since $\frac{\partial}{\partial\theta}M_a(0) = EX_a = a$ and $M_a(0) = 1$.

By (A.5) and differentiation we get

$$\frac{d}{da}M_a(\theta^*(a)) = \left(T_2 e^{\theta^*(a)T_2} - \int_{T_1}^{T_2} e^{\theta^*(a)x} F_a(x)\,dx\right)\frac{d}{da}\theta^*(a) - \theta^*(a)\frac{d}{da}\int_{T_1}^{T_2} e^{\theta^*(a)x} F_a(x)\,dx \qquad (A.10)$$

Evaluating (A.10) at $a = \bar{a}$ and using (A.6) we get

$$\frac{d}{da}M_{\bar{a}}(\theta^*(\bar{a})) = \bar{a}\frac{d}{da}\theta^*(\bar{a}). \qquad (A.11)$$

Note that $\mathcal{I}_a = \mathcal{I}_a(\bar{a})$ is the maximum of $f(a,\theta)$ over θ by (12) in Section 5. (A.3), (A.10), (A.11), and the fact that $M(\theta^*(\bar{a})) = M(0) = 1$ immediately imply

$$\frac{d}{da}\mathcal{I}_{\bar{a}} = \frac{d}{da}f(\bar{a},\theta^*(\bar{a})) = 0, \qquad (A.12)$$

Next we show that $\frac{d^2}{da^2}\mathcal{I}_{\bar{a}} > 0$. Using the fact that $M(\theta^*(\bar{a})) = 1$, it follows that

$$\frac{d^2}{da^2}f(\bar{a},\theta^*(\bar{a})) = \bar{a}\frac{d^2}{da^2}\theta^*(a) - \frac{d^2}{da^2}M_{\bar{a}}(\theta^*(\bar{a})) + \left(\frac{d}{da}M_{\bar{a}}(\theta^*(\bar{a}))\right)^2. \qquad (A.13)$$

Furthermore, by differentiating (A.10) one more time at $a = \bar{a}$ and by (A.6) and (A.7) we get

$$\frac{d^2}{da^2}M_{\bar{a}}(\theta^*(\bar{a})) = \bar{a}\frac{d^2}{da^2}\theta^*(\bar{a}) + \left(\frac{d}{da}\theta^*(\bar{a})\right)^2 E(X_{\bar{a}}^2) + 2\frac{d}{da}\theta^*(\bar{a}). \qquad (A.14)$$

Substituting (A.9), (A.11), (A.14) into (A.13) we get

$$\frac{d^2}{da^2}f(\bar{a},\theta^*(\bar{a})) = -\left(\frac{d}{da}\theta^*(\bar{a})\right)^2 \operatorname{var} X_{\bar{a}} - 2\frac{d}{da}\theta^*(\bar{a}) = \frac{1}{\operatorname{var} X_{\bar{a}}}, \qquad (A.15)$$

where the last equality follows from (A.9). Recall that the rate function $\mathcal{I}_a = (a,\theta^*(a))$. Thus, (A.15) implies $\frac{d^2}{da^2}\mathcal{I}_{\bar{a}} > 0$. Since by assumption, $M_a(\theta)$ is three times continuously differentiable in a neighborhood of $(\bar{a},0)$, (A.8) implies that $\theta^*(a)$ is twice continuously differentiable for all $a > \bar{a}$ which are sufficiently close to $\bar{a}$. Thus, $\mathcal{I}_a = f(a,\theta^*(a))$ is continuously differentiable

for all $a > \bar{a}$ which are sufficiently close to $\bar{a}$. We can therefore find a constant $k > 0$ such that

$$\frac{d^2}{da^2}\mathcal{I}_a > 2k > 0, \tag{A.16}$$

for all a in a neighborhood of $\bar{a}$. Thus, developing $\mathcal{I}_a$ in a Taylor series at $a = \bar{a}$ we get

$$\mathcal{I}_a = \mathcal{I}_{\bar{a}} + \frac{d}{da}\mathcal{I}_{\bar{a}}(a - \bar{a}) + \frac{d^2}{da^2}\mathcal{I}_{\tilde{a}}\frac{(\tilde{a} - \bar{a})^2}{2}, \tag{A.17}$$

for an $\tilde{a}$ between $\bar{a}$ and a. The Lemma now follows from (A.12), (A.16) and (A.17) since $\mathcal{I}_{\bar{a}} = 0$ by (A.4) and footnote 31.

We now use Lemma 3 to prove our main convergence result.

Proof of Proposition 1. Let $Z \geq \bar{z}$. By (13) in Section 5.1, the bank's probability of default is bounded above by

$$\int_{\bar{z}}^{\infty} e^{-\mathcal{I}_z(\bar{R}^*)n}\, dH(z).$$

Choose $z_1 > \bar{z}$. Since $\mathcal{I}_z(\bar{R}^*)$ is monotone increasing in z (cf., Stroock (1984, Lemma 3.3.) we get

$$\int_{z_1}^{\infty} e^{-\mathcal{I}_z(\bar{R}^*)n}\, dH(z) \leq e^{-\mathcal{I}_{z_1}(\bar{R}^*)n} P(\{Z \geq z_1\}). \tag{A.18}$$

Clearly, this expression converges to zero exponentially as $n \to \infty$, i.e., it can by bounded above by a term $e^{-k_2 n}$. It therefore remains to give an estimate for the rate of convergence if the realization of Z is between $\bar{z}$ and z_1. We can get such an estimate by using Lemma 3.[32] Furthermore, note that $E[R(Y_1 + z)]$ converges to $E[R(Y_1 + \bar{z})]$ at the same rate as $z \to \bar{z}$. This and Lemma 3 implies that $\mathcal{I}_z(\bar{R}^*) \geq k(z - \bar{z})^2$. It therefore remains to derive the rate of convergence of

$$\int_{\bar{z}}^{z_1} e^{-k(z-\bar{z})^2 n}\, dH(z) \tag{A.19}$$

[32]Let X_a be $R(Y_1 + z)$ where $a = E[R(Y_1 + z)]$, and let $\bar{a} = E[R(Y_1 + \bar{z}(\bar{R}^*))]$. Then all assumptions of Lemma 3 are fulfilled. Differentiability follows since the function $z \mapsto e^{R(Y_1+Z)}$ is differentiable except on a set of measure zero. Thus, Leibnitz's rule of differentiation under the integral can be applied.

as $n \to \infty$. Let h denote the density function for H. Substituting z by $\frac{z-\bar{z}}{\sqrt{n}}$ in (A.19) implies

$$\int_{\bar{z}}^{z_1} e^{-k(z-\bar{z})^2 n}\, dH(z) \leq \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} e^{-kz^2} h(\bar{z} + \frac{z}{\sqrt{n}})\, dz. \qquad (A.20)$$

Note that $e^{-z^2} h(\bar{z} + \frac{z}{\sqrt{n}})$ converges to $e^{-z^2} h(\bar{z})$ as $n \to \infty$. Thus, there exists a constant k_1 such that the right-hand side of (A.20) is bounded above by $\frac{k_1}{\sqrt{n}}$. The result follows now from (A.18) and (A.20).

References

P. Billingsley (1968), "Convergence of Probability Measures," John Wiley & Sons: New York.

D.W. Diamond (1984), "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies* **51**, 393–414.

K. Dowd (1991), "Optimal Financial Contracts," forthcoming *Oxford Economic Papers*.

D. Gale and M. Hellwig (1985), "Incentive–Compatible Debt Contracts: The One Period Problem," *Review of Economic Studies* **52**, 647–663.

M. Gertler (1988), "Financial Structure and Aggregate Economic Activity: An Overview," *Journal of Money, Credit and Banking* **20(3)**, 559–587.

S. Krasa and A.P. Villamil (1991a), "Monitoring the Monitor: an Incentive Structure for a Financial Intermediary," forthcoming *Journal of Economic Theory*.

S. Krasa and A.P. Villamil (1991b), "Optimal Contracts with Costly State Verification: the Multilateral Case," BEBR Working Paper 91–0127, University of Illinois.

J. Panzar (1989), "Technological Determinants of Firm and Industry Structure," Handbook of Industrial Organization I, Elsevier Science Publishers: Amsterdam.

K. Parthasarathy (1977), "Introduction to Probability and Measure," McMillan Press: London.

E.C. Prescott (1986), "Theory Ahead of Measurement," Quarterly Review, Minneapolis Federal Reserve Bank: Minneapolis, MN.

J. Stroock (1984), "An Introduction to the Theory of Large Deviations," Universitext, Springer.

R.M. Townsend (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* **21**, 1–29.

R.M. Townsend (1988), "Information Constrained Insurance: The Revelation Principle Extended," *Journal of Monetary Economics* **21**, 411-450.

S. Varadhan (1984), "Large Deviations and Applications," CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA.

S.D. Williamson (1986), "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing," *Journal of Monetary Economics* **18**, 159–179.

S.D. Williamson (1989), "Restrictions on Financial Intermediaries and Implications for Aggregate Fluctuations: Canada and the United States 1870-1913," *NBER Macro Annual* **4**, 303-340.

HECKMAN
BINDERY INC.
JUN 95
Bound-To-Please N. MANCHESTER,
INDIANA 46962

UNIVERSITY OF ILLINOIS-URBANA
3 0112 004996879